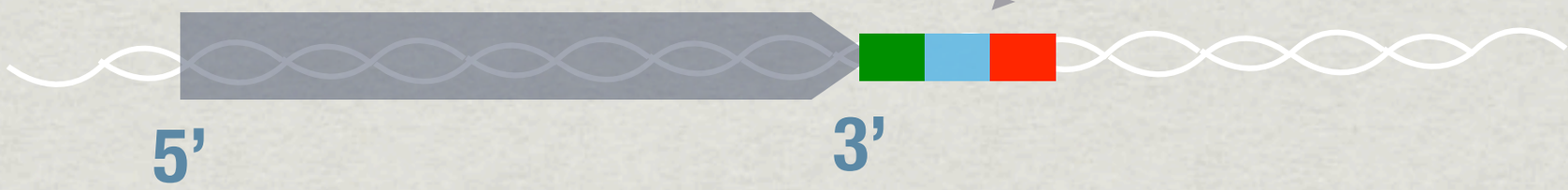
Experimental Techniques 2

- High-throughput interaction detection
- Yeast two-hybrid - pairwise
  - organisms as machines to learn about organisms
  - yeast, worm, fly, human,...
  - low intersection between repeated experiments
  - *in vivo*, but takes place inside the nucleus.
  - Estimated 50% FP rate
- TAP-MS (co-immunoprecipitation) - complexes

Tandem Affinity Purification (Puig et al, 2001)

Want to find interaction partners for protein encoded by this gene:

Add a tag to the end of its DNA sequence.



Calmodulin binding peptide

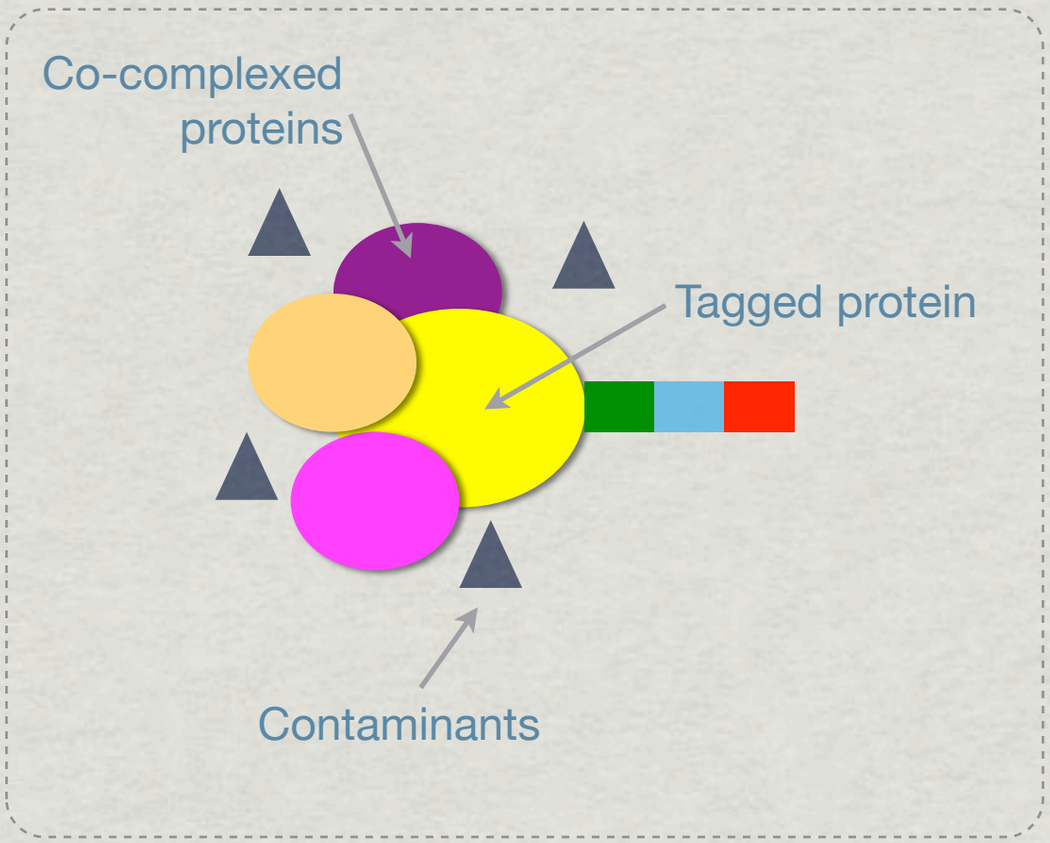
“Protein A” from *Staphylococcus aureus* Binds to IgG protein

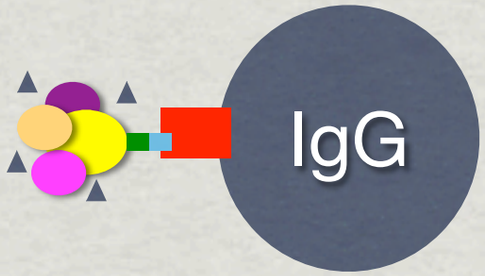
TOBACCO ETCH VIRUS PROTEASE

TEV protease cleavage (cutting) site

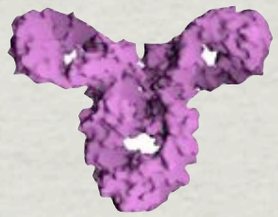
Glu-Asn-Leu-Tyr-Phe-Gln Gly

Fishing for Proteins

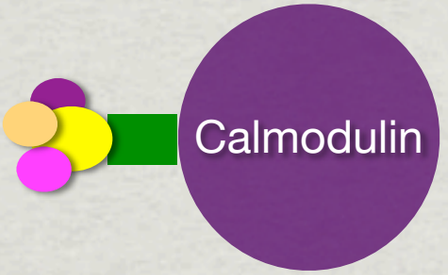




Grab with Immunoglobulin G protein



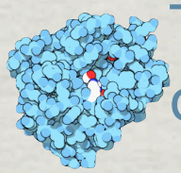
Wash contaminants cleave with TEV



Retrieve with calmodulin beads

- Tag may not be exposed
- Tag may change folding / binding properties
- Tag may change expression levels

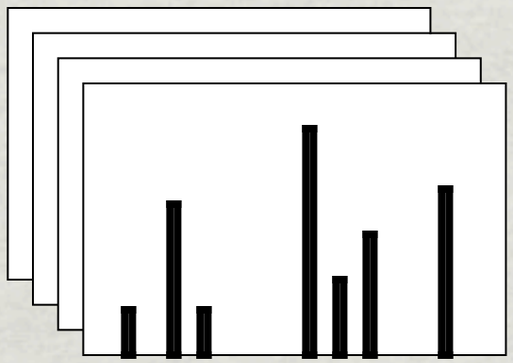
Sequencing Proteins (Tandem Mass Spectrometry)



Trypsin digestion

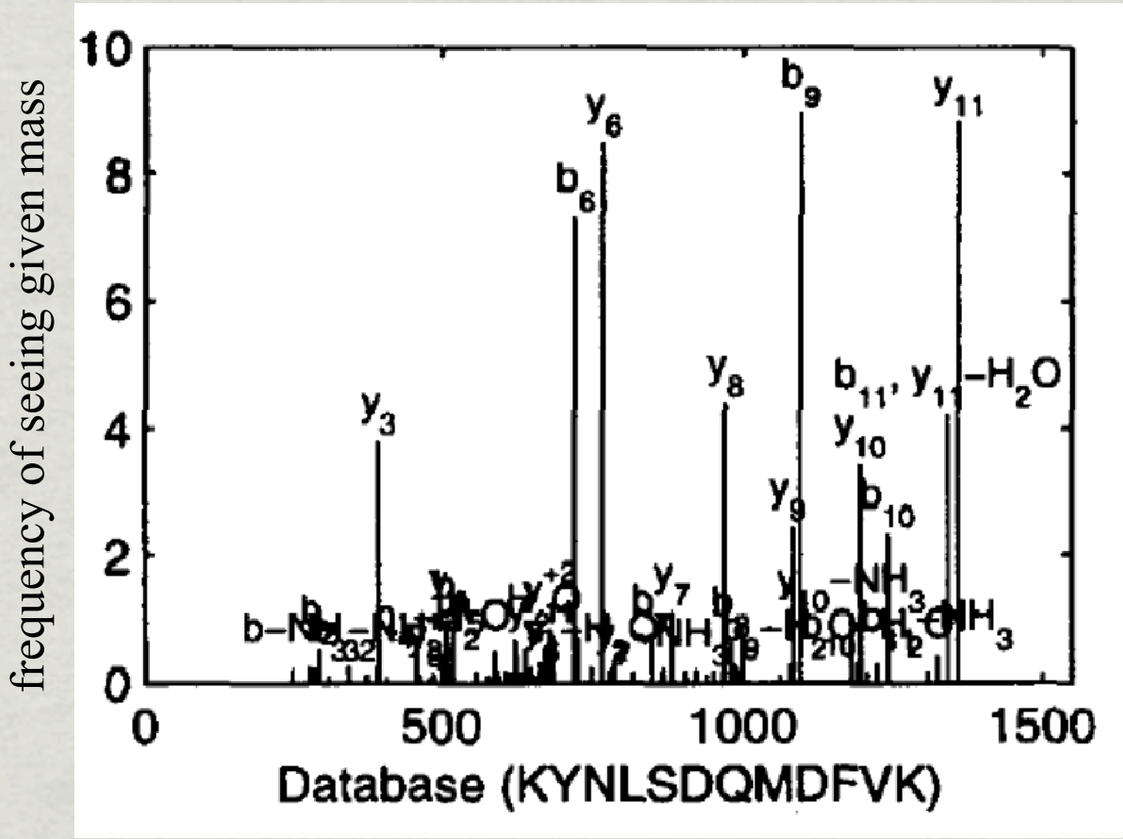
AAVEK → (Tandem Mass Spec)

b-ions	y-ions
A	K
AA	EK
AAV	VEK
AAVE	AVEK
AAVEK	



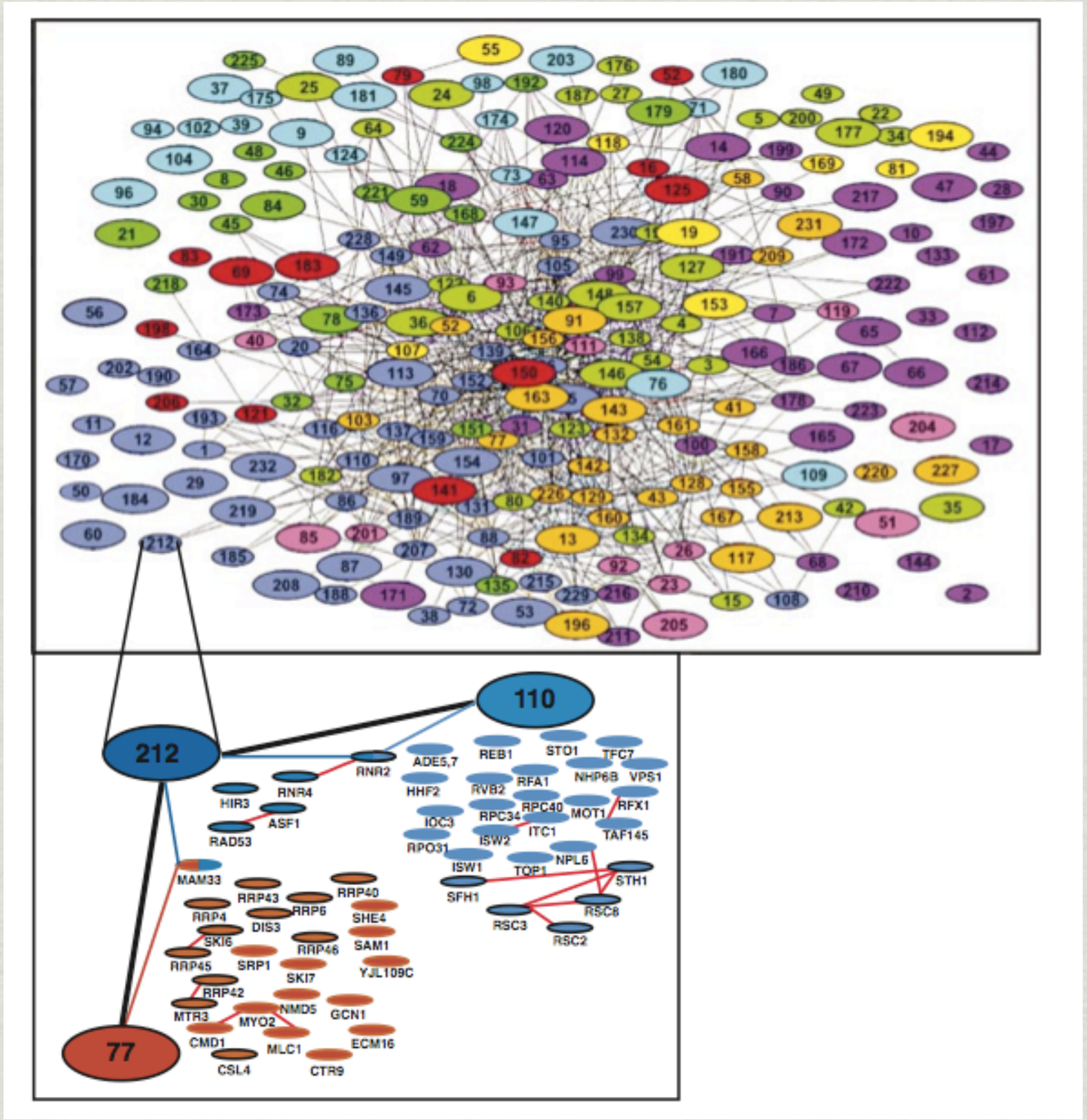
Search

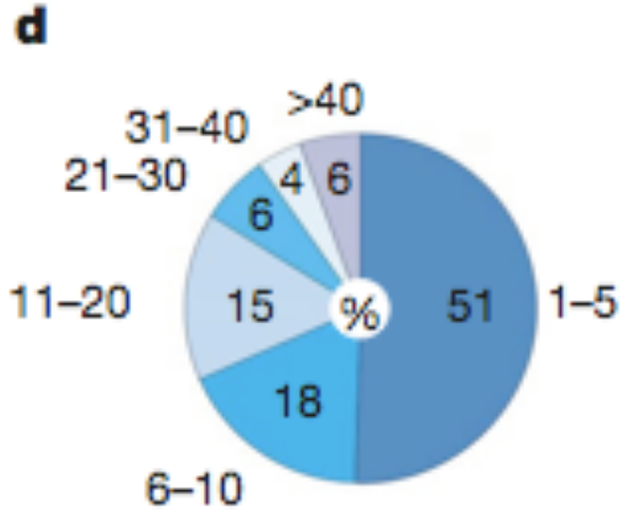
Database of known or predicted spectra



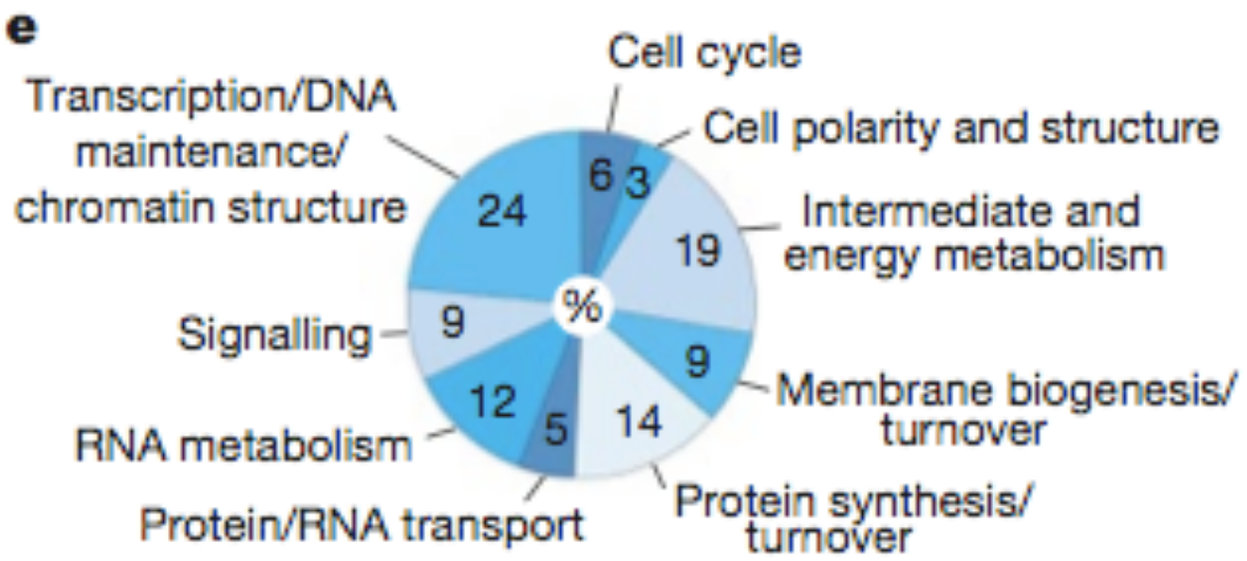
Gavin et al, 2002 Results:

- 589 tagged proteins
  (78% of which returned some interaction partners)
- 232 complexes
  (grouping those with substantial overlap)
- Covering 1440 proteins
- Not binary interactions
- In this picture: edges mean complexes share a protein





Number of proteins per complex

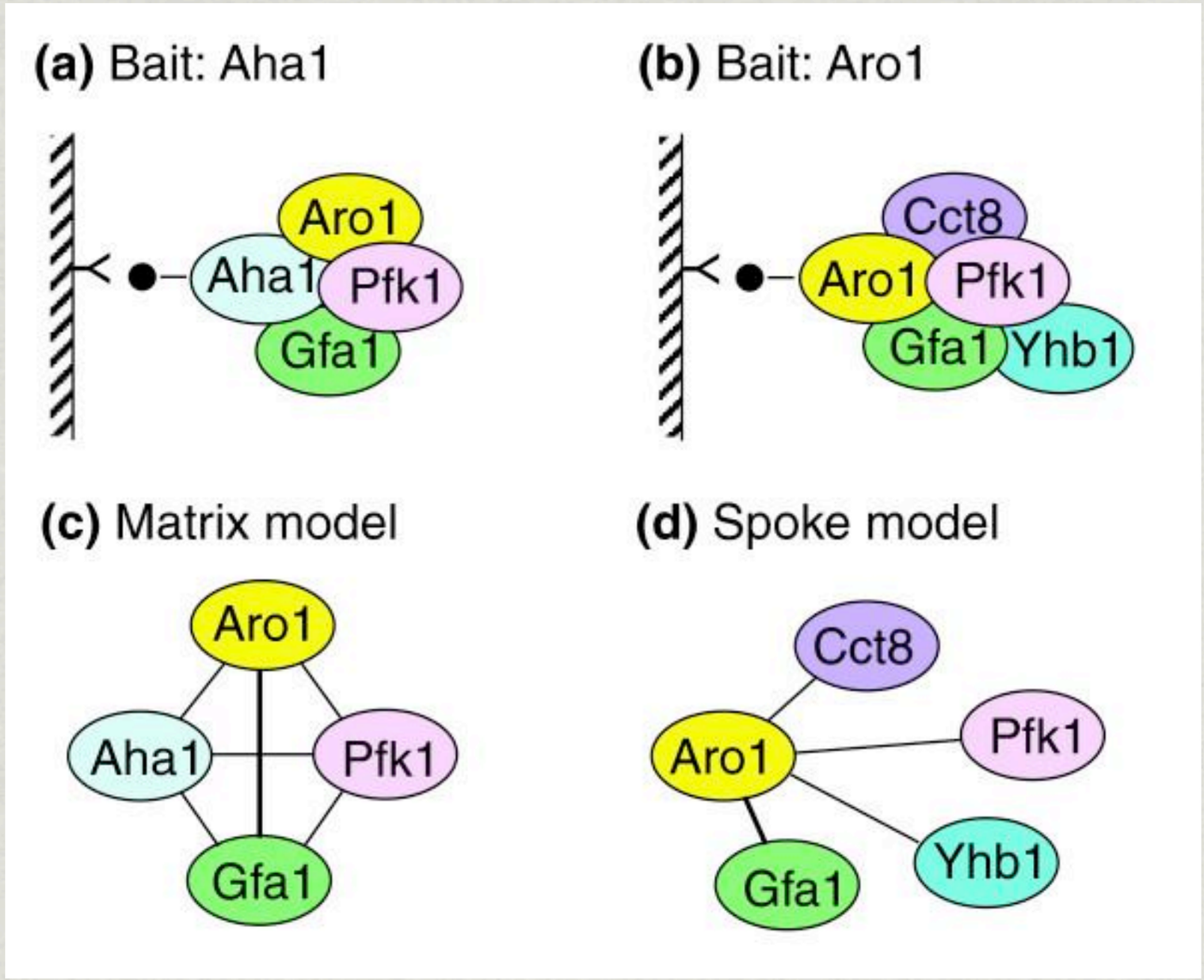


Distribution of complexes according to function

Gavin et al, 2006 - Larger scale TAP-MS:

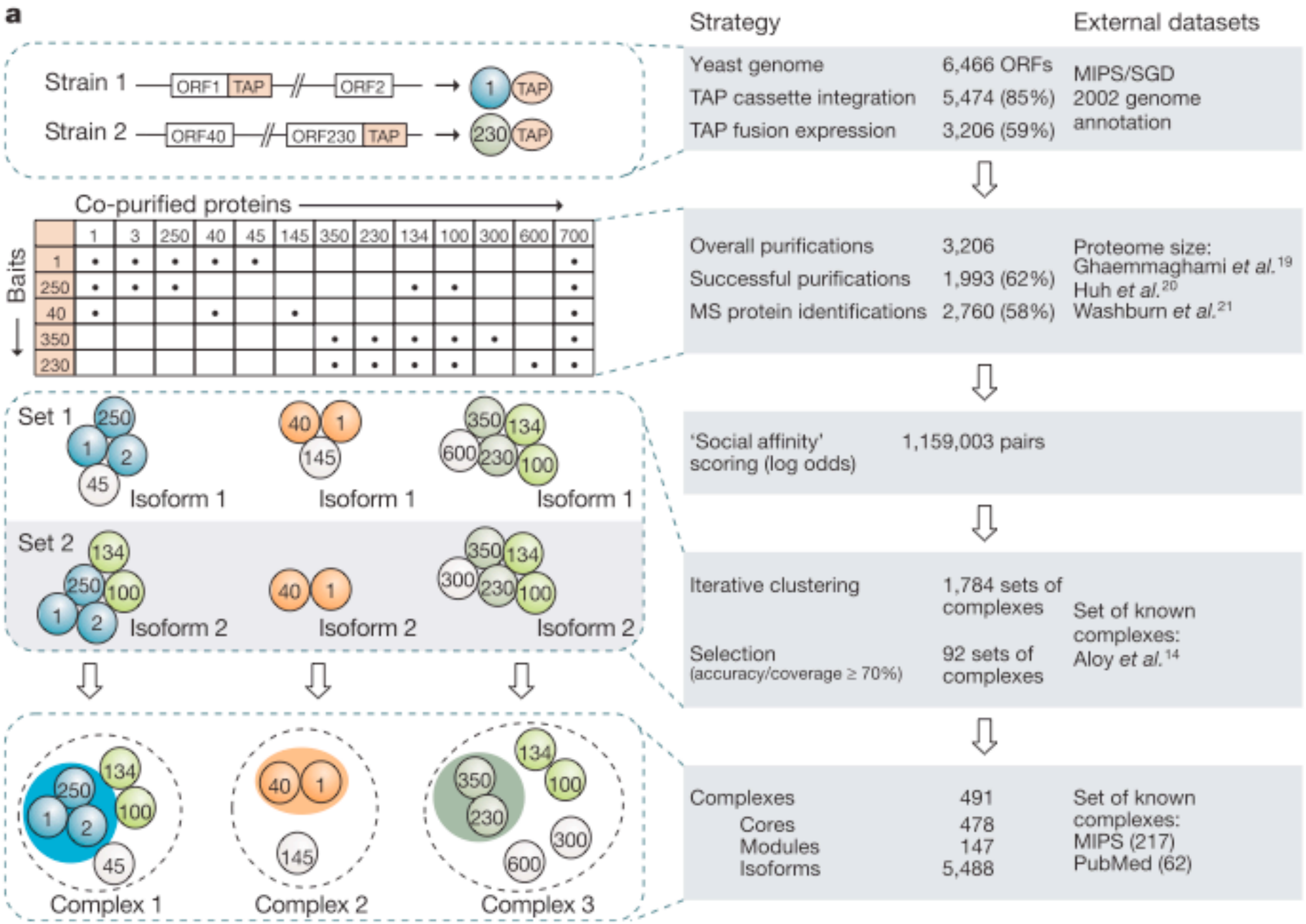
- 2006 update:
  - 2,760 unique proteins involved in some complex (60% of the proteome of yeast)
  - Reproducible: repeated experiment for 139 proteins, and 69% of retrieved proteins were common to both experiments.
  - 73% of the known complexes in MIPS (database) were found.
  - ~ 491 complexes (more about how this is defined later)
  - Of which 257 were novel

Simple ways to Convert to a Graph



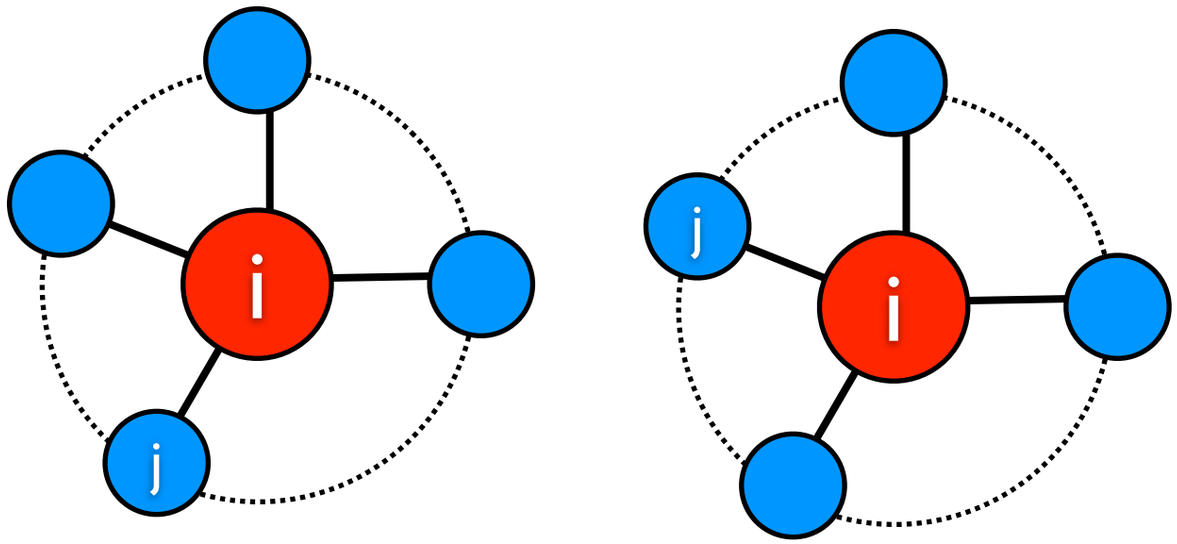
Goll & Uetz, 2006

Gavin et al, 2006 - Larger scale TAP-MS:

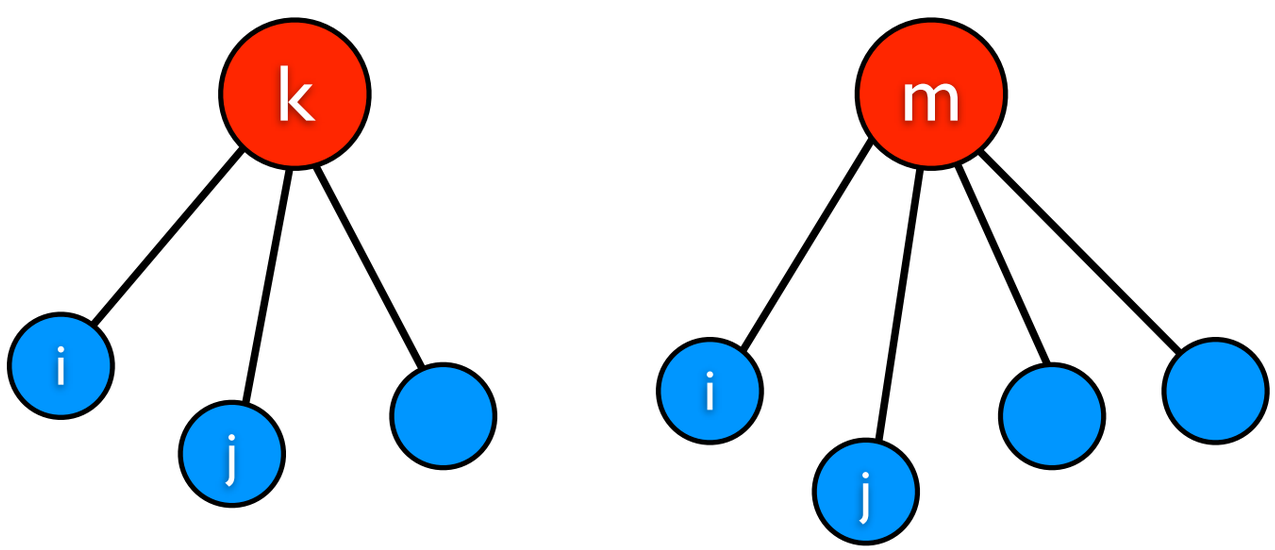


Socio-affinity Index

$A(i,j) := S_{i,j\,|\,i=bait} + S_{i,j\,|\,j=bait} + M_{i,j}$



$S_{i,j\,|\,i=bait} \approx$ ratio of # of times j was retrieved using i as bait, divided by the expected # of times, given how often j appears and how many preys i brings in.



$M_{ij} \approx$ ratio of # of times i and j both seen when using some other bait divided by the expected # of times, given how often i and j appear.

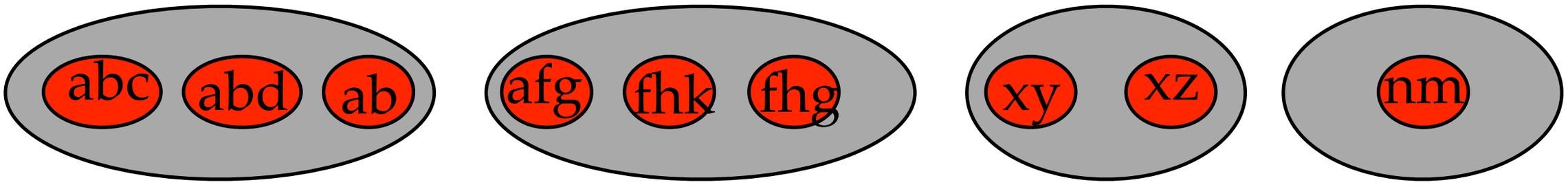
Clustering and Cluster Ensembles

- The clustering algorithm to find complexes:
  1. Using A(i,j) as a similarity metric, cluster the proteins (using some algorithm: UPGMA, single linkage, complete linkage).
  2. Use a threshold of similarity X to define clusters.
  3. **Subtract a penalty (e.g. 0.5, 1, or 2) from A(i,j) where i,j are in the same cluster and go to step 1.**
  4. Stop after between 2 and 10 iterations.

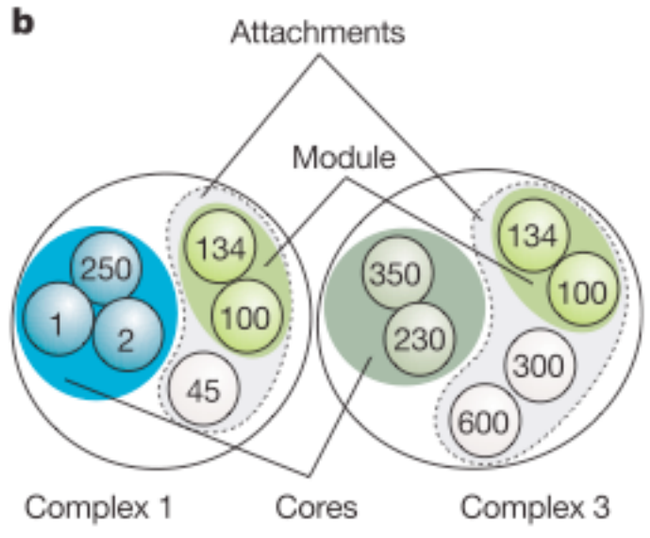
- Note: algorithm is underspecified. So: repeat with many different choices of parameters, take clusters found with a set of parameters that resulted in > 70% coverage and accuracy.

Isoforms & core and attachment proteins

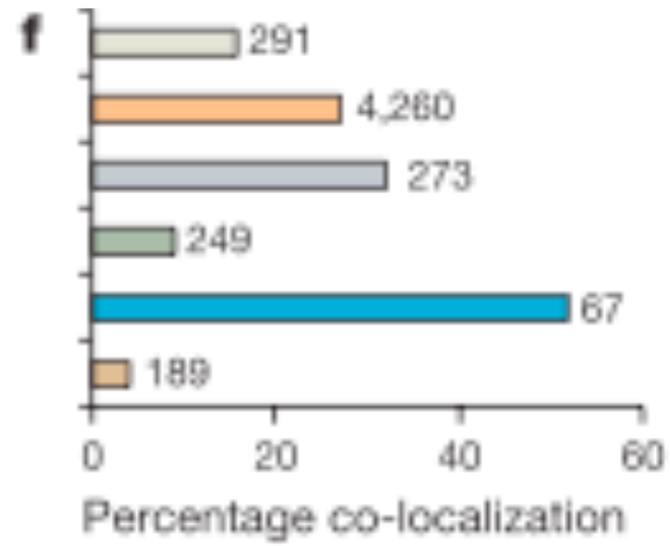
- 5,488 different clusters => “isoforms”
- Group together similar clusters into “complexes:”

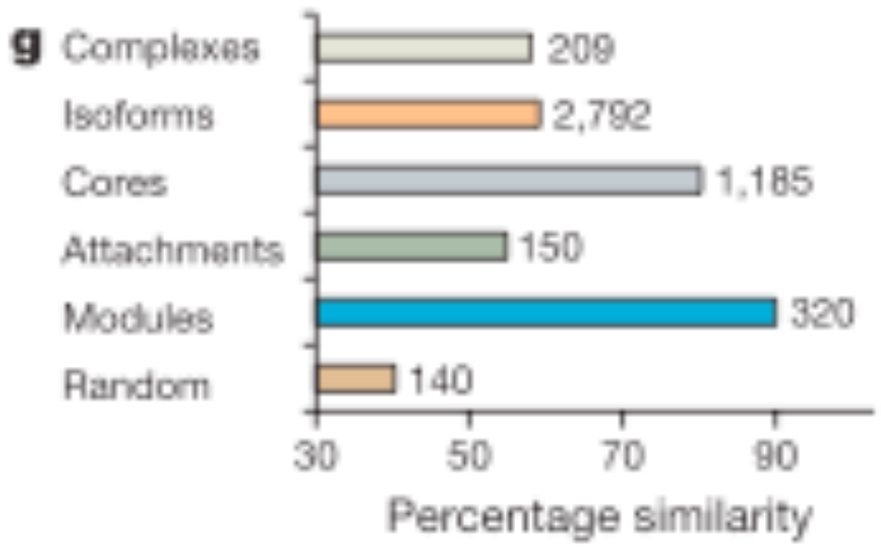


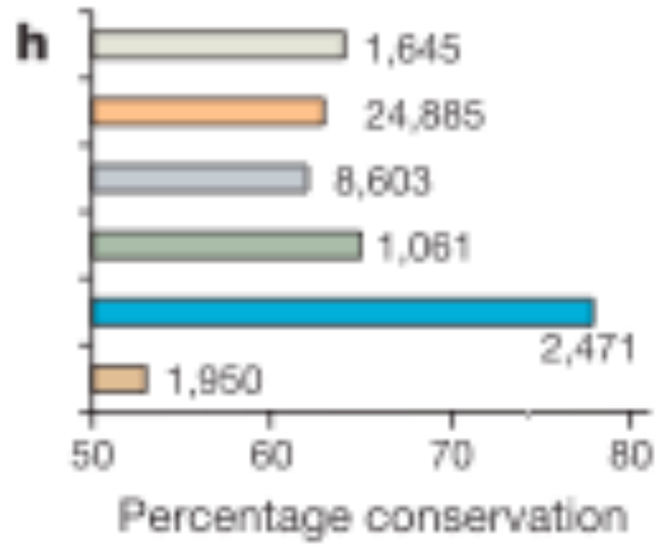
- Cores = subsets seen in most of the clusters within one group (average size 3.1 ± 2.5)
- Modules = pairs that were always together and seen in > 1 complex.
- Attachments = proteins not in the core.

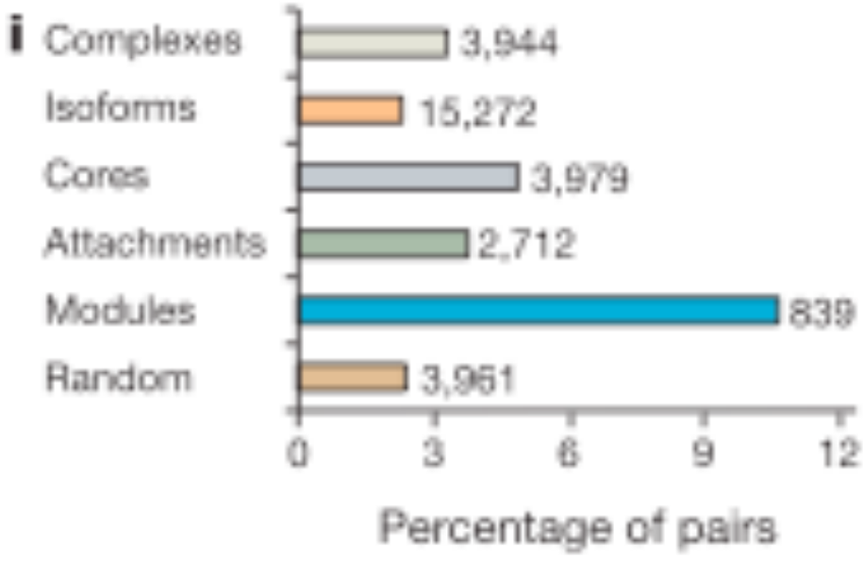


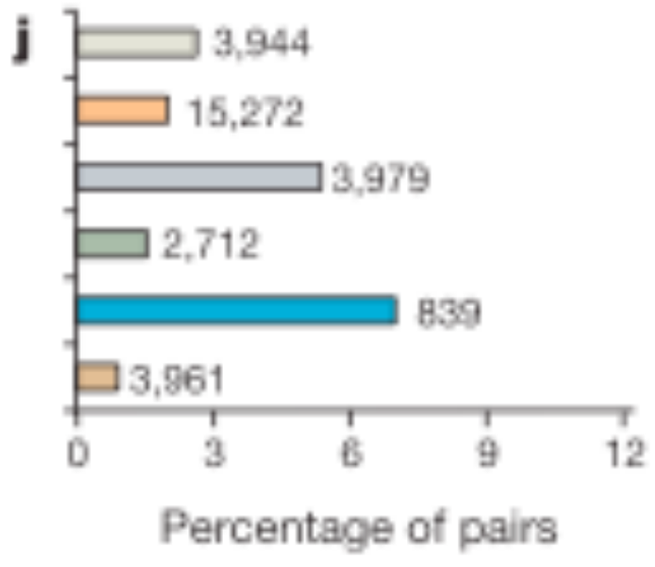
f, g, h: % of pairs co-localized, same cellular function, conservation.











% of pairs known from structures or Yeast 2 Hybrid

TAP-MS vs. Yeast 2 Hybrid

- Yeast 2-hybrid:
  - Pro: better at transient interactions (b/c they only have to happen long enough to “turn on” the reporter gene)
  - Con: take place in nucleus (may be unnatural)
  - Con: only binary interactions

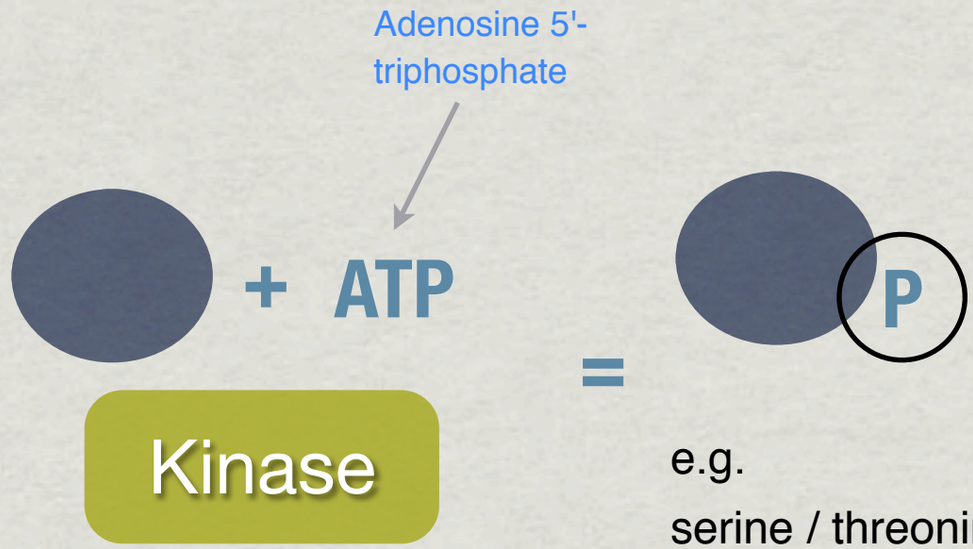
- TAP-MS:
  - Pro: can find higher-order interactions (> binary)
  - Con: requires more stable interactions

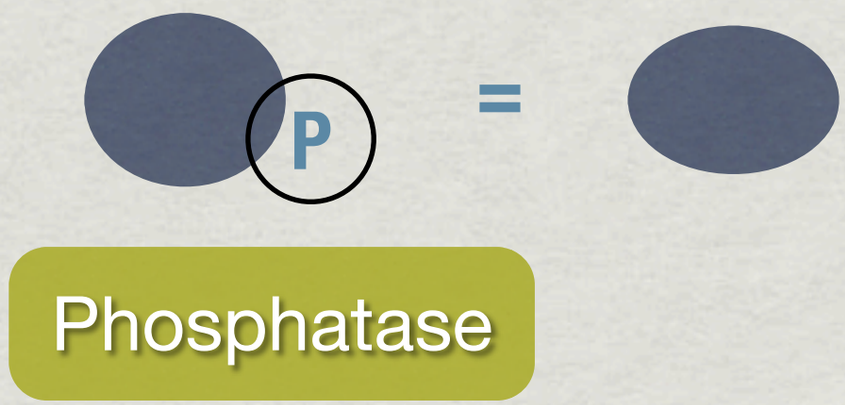
Ho et al, 2002 Results:

725 yeast proteins chosen to be “bait”:

#	Protein Function
100	Kinases
36	Phosphatases
86	DNA damage response
503	Other proteins

600 baits worked (~10% of yeast proteins)
493 specific baits
1,578 proteins involved in ≥ 1 interaction
3,617 interactions





Kinases / Phosphatases

kinase: class of enzyme (protein) that adds a phosphate group to other molecules (usually a protein).

phosphorylation: the process of adding a phosphate group (PO_4) to a protein.

Phosphorylation often changes the shape (conformation) of a protein, thereby turning it "on" or "off".

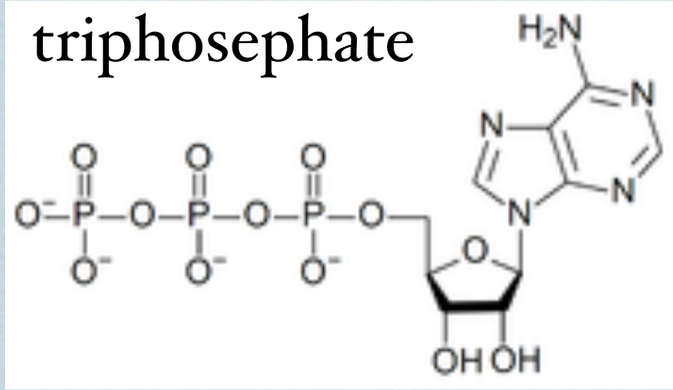
For example, phosphoylation can make a hydrophobic residue hydrophilic.

It is an important regulatory mechanism.

Estimate: >30% of proteins are phosphorylated in humans

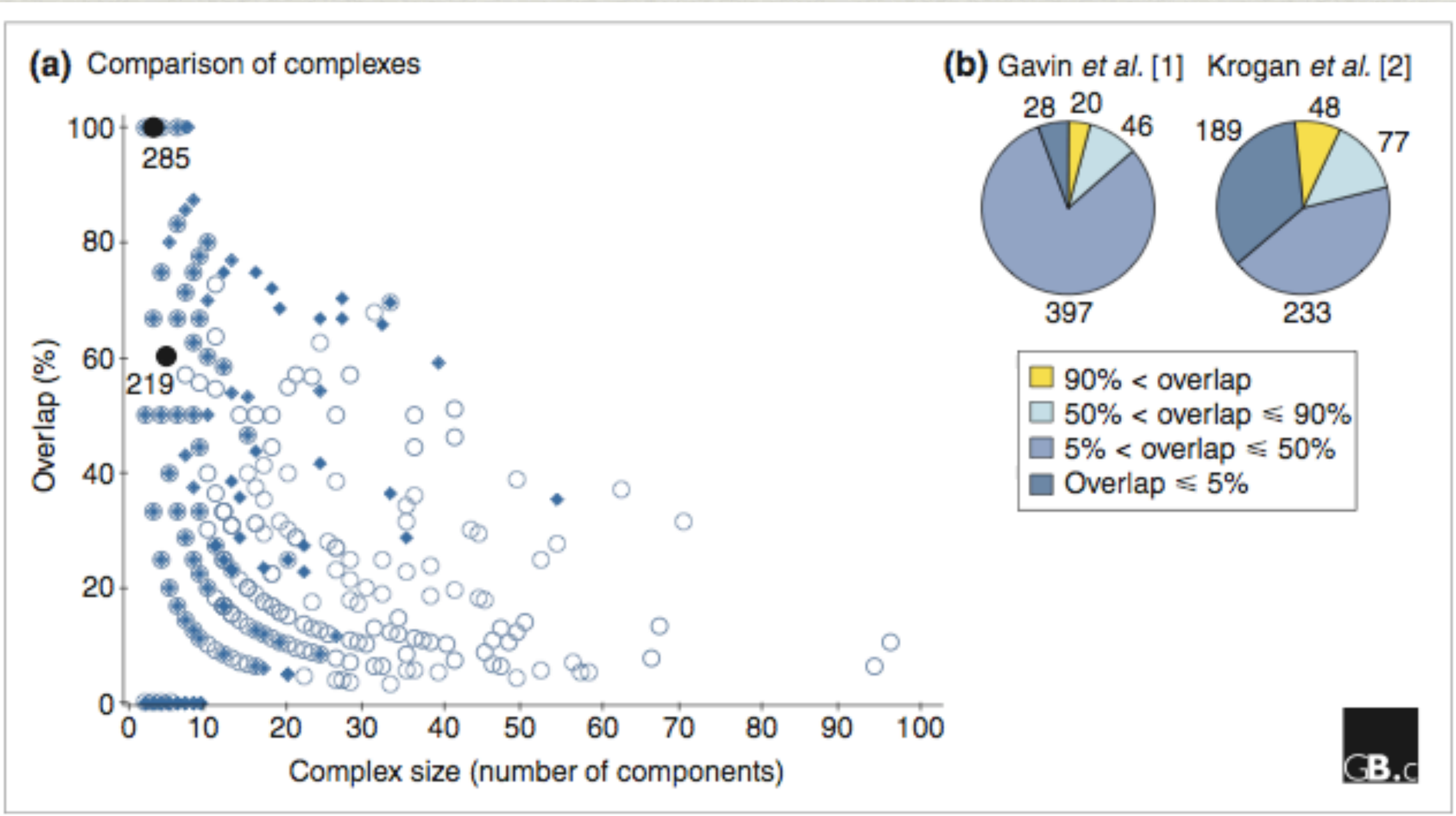
- 518 known kinases in human
- 122 known kinases in yeast

ATP: adenosine-5'-triphosephate



Comparing TAP Experiments

Goll & Uetz, 2006

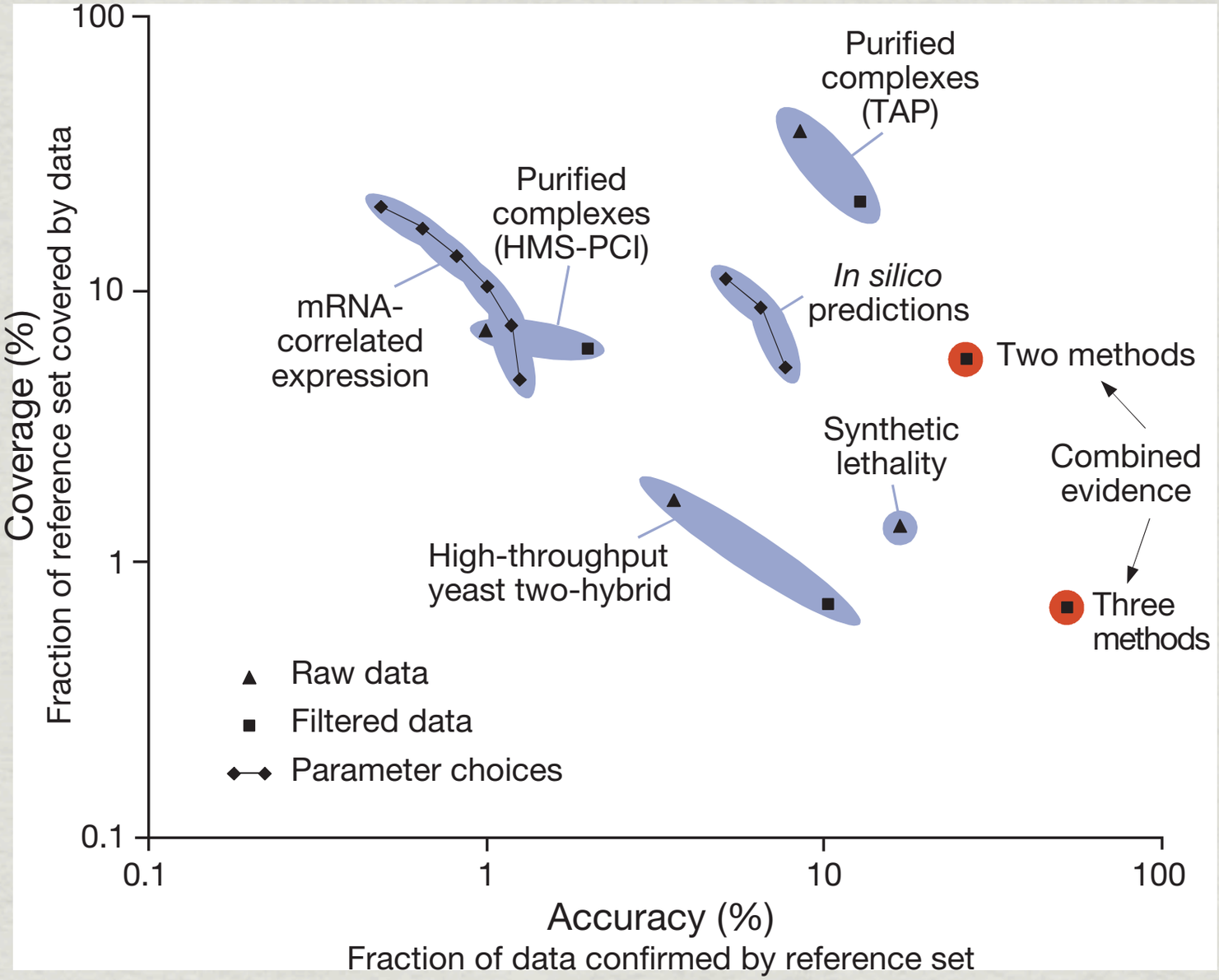


Von Mering et al, 2002 Comparisons

10,907 “trusted” interactions from YPD & MIPS (T)

Coverage = % of T also in D

Accuracy = % of D also in T

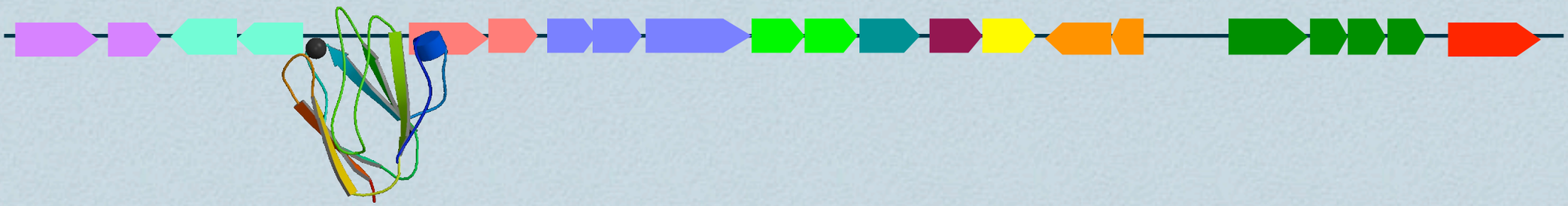


Combining methods again helps significantly. (But of 80,000 predicted interactions, only 2,400 were seen in more than 1 method.)

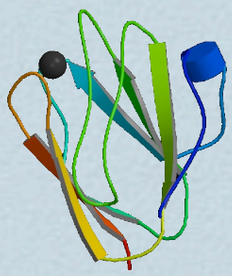
Von Mering estimate for # of interactions in yeast

- M = interactions seen more than once (2,400)
- 1/3 of them were previously known
- At the time: ~ 10,000 interactions known
- Therefore, expect 30,000 interactions total
- (Sprinzak et al estimate ~ 16,000)

Transcription network, aka regulatory network:



Transcription Factors =
proteins that bind to DNA
to activate or repress the
nearby, downstream genes.



regulates

gene

gene

the regulated gene might
also be a transcription factor

leads to a
directed graph

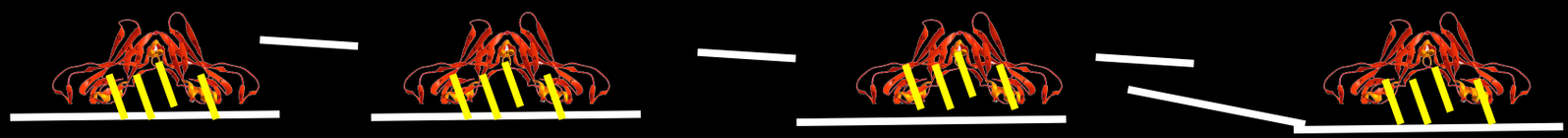
ChIP-chip (ChIP-seq)

Chromatin immunoprecipitation - chip

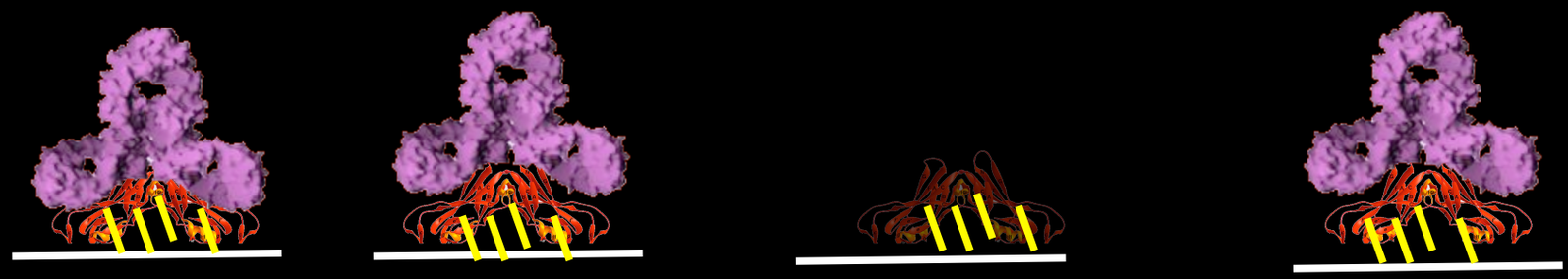
TF Binds to DNA

TF Cross-linked to DNA (covalent bonds)

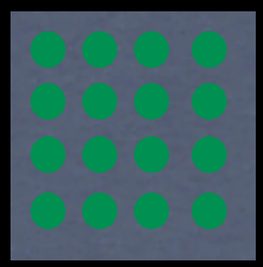
Cell is lysed, DNA fragmented



Antibodies used to pull out protein-DNA complexes



DNA is “read” using microarray or short-read sequencing



Synthetic Lethality

- Predicts a particular kind of functional interaction (“genetic interactions”)
- “Synthetic” b/c manufactured mutations

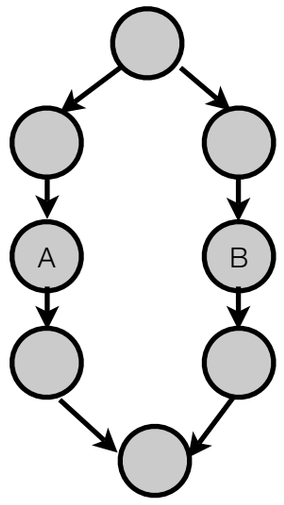
-A = survive

-B = survive

-A & -B = die

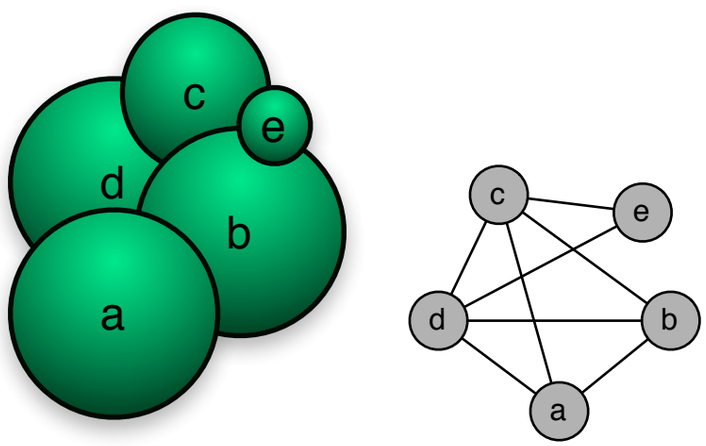
pretty course measurement

proteins A and B are likely to be involved in similar functions

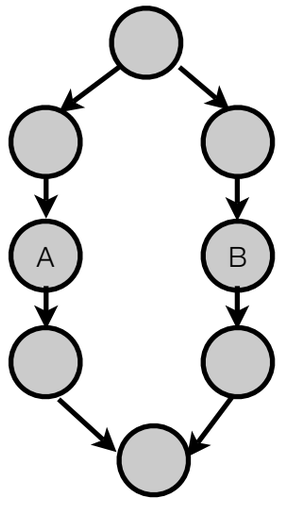


A & B are “redundant” or complementary (parallel pathways)

Explanations



Complex abcde can function when a single one of its proteins is removed, but not if 2 are removed.

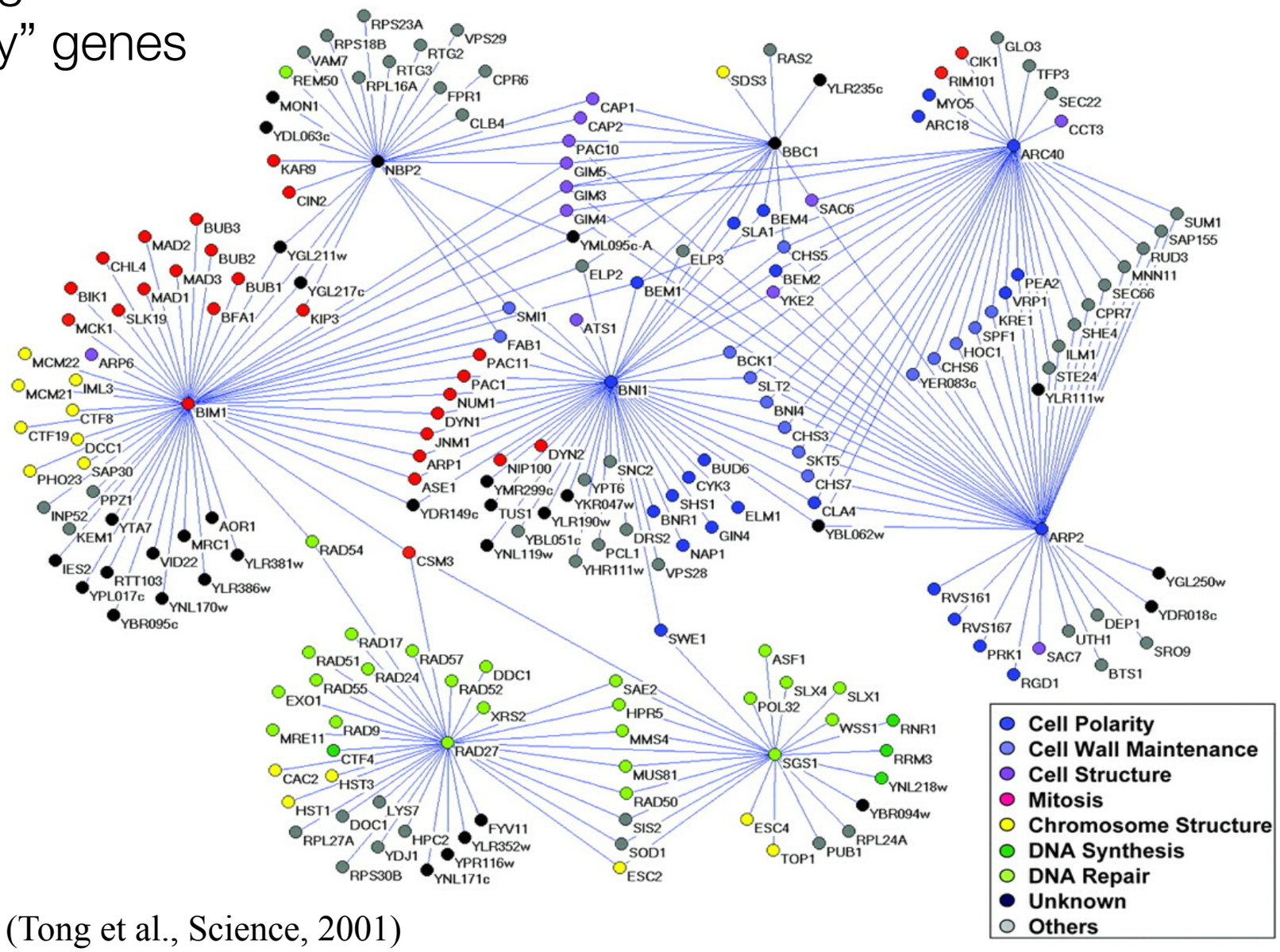


A & B are “redundant” or complementary (parallel pathways)

- Two copies of the same protein.
- Complexes that can function without one of their constituent proteins.
- **Two “redundant” pathways.**
- **3 pathways, where any 2 are required**

SSL network from 2001

8 query genes
4500 “array” genes



(Tong et al., Science, 2001)